Available online at www.sciencedirect.com

ScienceDirect

Genomics 91 (2008) 219–231

GENOMICS

www.elsevier.com/locate/ygeno

Review

Integrated approaches to uncovering transcription regulatory networks in mammalian cells

Kai Tan [a,*], Jesper Tegner [b,c,*], Timothy Ravasi [a,d,*]

[a] *Department of Bioengineering, Jacobs School of Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[b] *Unit of Computational Medicine, King Gustaf V Research Institute, Department of Medicine, Karolinska Institute, SE-171 76 Stockholm, Sweden*
[c] *Computational Biology, Department of Physics, Linköping University, SE-581 53 Linköping, Sweden*
[d] *Scripps NeuroAIDS Preclinical Studies Center, Scripps Research Institute, La Jolla, CA 92037, USA*

Received 17 September 2007; accepted 16 November 2007
Available online 8 January 2008

Abstract

Integrative systems biology has emerged as an exciting research approach in molecular biology and functional genomics that involves the integration of genomics, proteomics, and metabolomics datasets. These endeavors establish a systematic paradigm by which to interrogate, model, and iteratively refine our knowledge of the regulatory events within a cell. Here we review the latest technologies available to collect high-throughput measurements of a cellular state as well as the most successful methods for the integration and interrogation of these measurements. In particular we will focus on methods available to infer transcription regulatory networks in mammals.
Published by Elsevier Inc.

Keywords: Systems biology; Networks; Gene networks; Transcriptional networks; Genomics; Dynamics; Gene expression; Cellular networks; Personalized medicine; Systems medicine; Bioinformatics; Protein-protein interaction networks; ChIp-chip; Protein interaction maps

Contents

* Corresponding authors. J. Tegner is to be contacted at Unit of Computational Medicine, King Gustaf V Research Institute, Department of Medicine, Karolinska Institute, SE-171 76 Stockholm, Sweden. T. Ravasi, Department of Bioengineering, Jacobs School of Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Fax: +1 858 822 4246.
E-mail addresses: ktan@bioeng.ucsd.edu (K. Tan), jesper@karlavagen90.se (J. Tegner), travasi@bioeng.ucsd.edu (T. Ravasi).

0888-7543/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.ygeno.2007.11.005

Introduction

The genome content of most cells in a multicellular organism is identical. Nevertheless, in the human body over 220 different cell types can be distinguished based on their morphological properties, localizations, functions, and developmental stages [1,2].

The cell's ability to acquire and maintain its identity requires a tight control of the temporal and spatial interplay between millions of individual components (nucleic acids, proteins, and metabolites) encoded by the same genome [3–6]. Although biologists commonly study cells as if they were static entities, individual cells rarely operate in a steady state. The ability of a cell to detect and respond to transient signals within and outside the cell is crucial for the integrity of a multicellular organism. When error occurs in the fine-tuned cellular regulatory system or errors are introduced into the genome as mutations, the cellular behaviors and identities can deviate from their normal situations and thereby induce pathological scenarios such as diseases [7–9]. To understand these cellular functions in detail, we need to identify the mechanisms by which genomic information is tightly controlled in space and time.

In eukaryotes the control of gene expression is a multilevel process influenced by the position of each chromosome in the nucleus (nuclear territories) [10–12], the spatial and conformational rearrangements of chromosomal regions (chromatin remodeling), and the fine interplay between regulatory proteins and DNA sequences [11–13] (Fig. 1). The interactions between regulatory proteins and DNA sequences are probably the most diverse and complex of all these regulatory mechanisms. In any given cell all combinations of transcriptional regulators and their specific DNA binding sequences responsible for the correct regulation of every transcript encoded in the genome are usually referred to as the transcriptional regulatory network (TRN) of the cell. Yet recent findings add an additional dimension to the regulatory control of gene expression, in that evidence has been gathered [14] supporting the notion that transcription is itself controlled by noncoding RNAs (ncRNAs). The precise mechanism by which this posttranscriptional control operates may be very complex. For example, RNA may act directly on the induction, processing, or stability of another transcript. Furthermore ncRNAs are a major, regulated output of the mammalian genome [15,16]. The role of ncRNAs in mammalian transcription regulation is beyond the scope of this review, although they need to be considered as a part of the regulatory network.

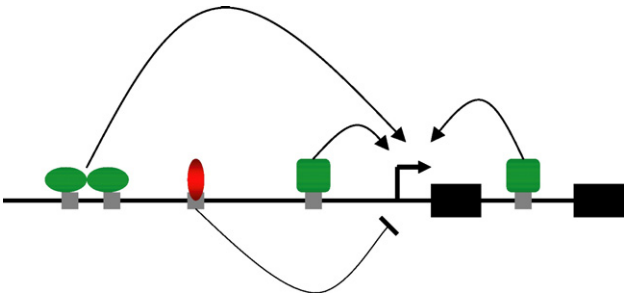

Fig. 1. Transcription regulation. Regulation of gene expression is a complex interplay between transcription regulators, transcription regulator complexes, and DNA *cis*-regulatory elements. Black line, DNA; black box, exon; black arrow, transcription start site; gray box, DNA *cis*-regulatory element; green symbol, transcriptional activator; red symbol, transcriptional repressor.

If we know the identity of every transcript produced by the genome in each cell type, the way transcription changes with time and space, and, most importantly, the control mechanisms of these genome outputs, the stage is set to infer the regulatory network for each cell under different conditions. This gives us the opportunity to suggest therapeutic mechanisms to correct for the errors underlying pathological conditions. Thus research aimed at developing techniques for uncovering cellular networks from experimental data is central for translating genomic discoveries into therapeutics.

During the past 5 years there has been rapid progress in the development of a systems approach for identifying transcriptional regulatory networks from high-throughput data [17,18]. A particularly powerful paradigm is the "perturbation" approach. Here the cellular response to a perturbation (typically of a regulator via environmental and/or genetic means) is monitored by high-throughput assays such as gene expression microarray, and fitting the data to a computational model of the gene network enables network identification provided large enough samples are available. Perturbation algorithms have been successfully applied to *Escherichia coli* and *Saccharomyces cerevisiae* [19,20]. Although successful in unicellular organisms, the perturbation approach has been difficult to apply to mammalian cells due to the limited number of samples and the difficulties in modulating gene activity in mammalian cells with siRNA or gene knockout. However, a complementary approach, data integration, promises to be a useful strategy for uncovering mammalian networks. In this paper we will therefore review the current advances in the construction of transcription regulatory networks in mammals. We will first focus on the available experimental and computational techniques to collect genome-wide measurements and to construct static models of regulatory networks. We will then describe some of the current approaches to add temporal and spatial dimensions to these networks.

Graphic theoretical representation of transcriptional regulatory networks

A graph is useful for illustrating a network, with its components and their interactions, of a complex system [21–26]. Here, the graph corresponding to the network is composed of "nodes," which can represent any biomolecules such as proteins, DNA, RNA, and metabolites, and "edges," representing relationships between nodes. An edge can denote a physical interaction, such as

protein–protein, protein–DNA, protein–RNA, RNA–RNA, and protein–metabolite interactions, or a functional relationship such as coexpression or a genetic interaction such as epistasis.

In a transcriptional regulatory network the nodes generally are transcription factors, cofactors, and chromatin regulators (from now on in this review we will refer to these genes collectively as transcription regulators or TRs) and the DNA regulatory regions (promoters and enhancers, referred as regulatory elements, REs). Edges in the network are physical interactions between TRs (in this case, protein–protein interactions) and physical interactions between TRs and REs (protein–DNA interactions, PDIs). Fig. 2 shows a graphical representation of a simple transcription regulatory network.

Experimental technologies to obtain network data

There are two main types of information we have to collect to be able to infer transcription regulatory networks in a cell. First, we need to identify all the nodes (including the TRs), which means that we have to define all the expressed transcripts under any developmental and growth conditions. Second, we need to identify all possible physical interactions between nodes, i.e., between TRs and between TRs and REs.

The completion of a large number of eukaryotic genome sequences, the rapid progress toward complete catalogs of genomic transcriptional output (the transcriptome), the capacity to analyze the complete transcriptome of a cell, and the advent of technologies that allow us to capture protein–protein and protein–DNA interactions within any cellular content offer us the tools to identify all the nodes and edges in the network.

Capturing the cellular transcriptome

Gene expression profiling using microarray chips is by far the most successful genome-wide technology to capture the genome output of a cell [27,28]. Expression microarray is an RNA-based method that allows the simultaneous measurement of virtually all the transcripts in a cell. This has been and still is a very powerful technique thanks to its relative technical simplicity, low cost, and short turnover time, which make expression microarrays a standard molecular biology technique available to any laboratory. Moreover, different array-based technologies have now become comparable across different platforms [29]. In recent years the advent of high-density microarray chips has allowed us to define the entire transcriptome of more complex organisms such as human. Computational methods used for the analysis of these large collections of data have also been improved and standardized, making the interpretation of microarray data more accessible to those without a strong computational background [30,31]. For these reasons and also thanks to the efforts of consortia aiming to standardize microarray datasets [32], we now have access to more then 160,000 expression profiles for various organisms and cell types (estimate from NCBI's Gene Expression Omnibus, July 2007; http://www.ncbi.nlm.nih.gov/geo).

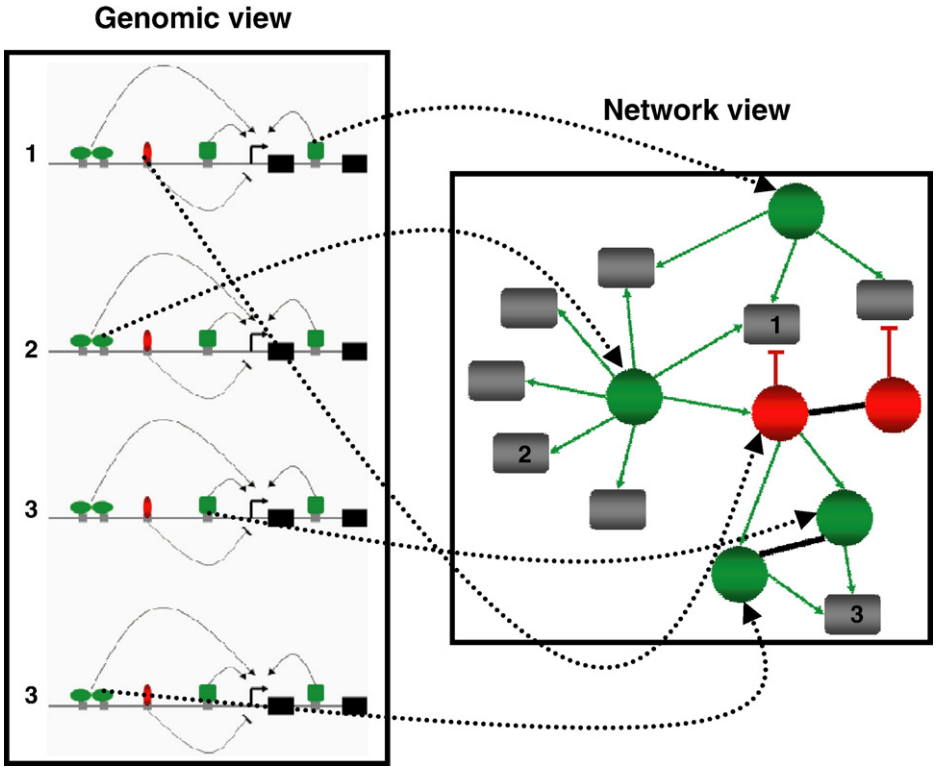

Fig. 2. Transcription regulatory networks. Regulatory events at the gene level (left) are mapped into a network view (broken arrows) aiming to capture all the regulatory events inside a cell (right). Protein–protein (black edges) and protein–DNA interactions (green and red edges) involving TRs (circular nodes) and REs (usually represented as the regulated genes, squared nodes) are shown. Note that protein–DNA interactions are directional and can represent activation (green) or repression (red).

Although transcript profiling using microarrays is the most widespread functional genomics technique, it is not the only one. Other approaches have been developed as alternatives to microarray, including serial analysis of gene expression (SAGE) [33,34] and most recently the cap analysis of gene expression (CAGE) [35–37]. More sensitive techniques such as the polony multiplex analysis of gene expression (PMAGE) allow the measurements of mRNAs as rare as one transcript per three cells [38]. Since SAGE and CAGE are RNA library based they require sequencing of millions of cDNA tags from each library and consequently they are not accessible to many laboratories. Unlike SAGE, CAGE also enables systematic and genome-wide mapping of transcription starting sites (TSSs) for every transcript expressed in a cell (more about the properties of CAGE is given in the following paragraphs).

Although with less throughput than array-based technologies, quantitative real-time PCR (qRT-PCR) is becoming an increasingly important complementary tool for the construction of TRNs [39,40], due to its quantitative nature and higher sensitivity, which allow more accurate measurements of low-abundance transcripts such as those encoding transcription factors [41] (Table 1).

Acquiring physical interaction data of the network

In a transcriptional regulatory network there are two types of physical interactions represented by edges, namely those between the regulatory proteins and their DNA binding sequences (PDIs) and those between regulatory proteins (PPIs).

Protein–protein interaction networks

In eukaryotes the regulation of gene expression often requires more than one TR to ensure the correct expression of a gene. TRs interact to form protein complexes and in many cases this is a requirement to be able to bind DNA regulatory elements [42–48]. For example, this is the case for homodimers binding palindromic transcription factor binding sites (TFBS) [49]. Furthermore, TFBSs tend to cluster together in specific and conserved regions in the genome and TRs targeting these DNA regions also interact at the protein level to form protein complexes [50–53]. A dramatic example is the transcriptional initiation complex, which is composed of more than 30 proteins and binds specific regulatory elements via a few core components such as the TATA box binding protein [54–58].

The interplay between TRs is often referred to as the combinatorial regulation of gene expression. Therefore capturing all possible combinatorial interactions between TRs is an essential step toward the construction of mammalian transcription regulatory networks. For this purpose complete maps of PPIs are of utmost value as a first step in mapping putative pairwise interactions.

PPIs are usually generated by two-hybrid technologies (Y2H) [59], in which two proteins, a bait fused with a specific DNA binding domain able to bind the yeast GAL4 gene and a prey fused with the GAL4 activator domain, are overexpressed in the same yeast cell. If the two proteins of interest are able to interact, then the GAL4 activator is reformed and thus able to activate the expression of a reporter gene that is under the control of the GAL4 gene promoter [60]. A similar approach has been developed, using a mammalian system, by Suzuki and colleagues [61,62]. Their technique is at least in principle more amenable to a systematic screening of binary mammalian PPIs because the overexpressed proteins are folded and modified in a more natural environment.

PPI maps can also be constructed using coimmunoprecipitation followed by mass spectrometry [63–69]. This technology is more specific than Y2H (lower false positive rate) and therefore less scalable. Since the technology relies on coimmunoprecipitation it is more suitable for identifying protein complexes with indirect interactions, in contrast to Y2H, which measures direct pairwise interactions.

In recent years the number of binary nonredundant human PPIs has increased dramatically thanks to extensive literature mining (36,617 in the HPRD database [70]) and also to large-scale experimentally determined PPIs such as the work from Rual and colleagues and Ewing and colleagues [63,71]. However, one of the limitations with the current human PPI map is the low coverage of TR interactions because the experimental techniques generally are biased toward large macromolecular complexes (i.e., ribosome, spliceosome, membrane channels, etc.) and because of the low abundance of TRs compared to cytosolic proteins. Suzuki and colleagues of the RIKEN Genome Science Center in Japan have generated for the first time a nuclear-specific PPI map for mouse [62] and now they are focusing on the human nuclear PPI map (personal communication). Such maps will be very useful resources for the construction of mammalian transcriptional regulatory networks.

Another limitation of current PPI networks is the extremely high false discovery rate, which has been estimated to be from 40 to 70% for the Rual human PPI network [71,72]. Such high error rates can be readily reduced with the integration of other types of data, for example, gene expression information, so that all those interacting proteins that are never coexpressed could be removed from the network. We will discuss more details about the integrated approach in the following sections (Table 1).

Table 1
Experimental techniques commonly used to collect high-throughput measurements of gene expression and physical interactions

Technique	References
Expression measurement	
DNA microarray	[27,28]
SAGE	[33,34]
CAGE	[35–37]
PMAGE	[38]
qRT-PCR	[39,40]
Protein–DNA measurement	
ChIP-chip	[74,75]
ChIP-PET	[76]
Protein–protein measurement	
Two-hybrid systems	[59,61,62]
Co-IP and mass spectrometry	[63–65]

Protein–DNA interaction networks

To regulate gene expression, either individual TRs or complexes of TRs need first to bind specifically to *cis*-regulatory DNA sequences. Traditionally, the most common methods to infer TR–DNA binding events on a large scale are computational ones (discussed in detail later).

New technologies have emerged that enable in vivo genome-wide experimental mapping of TR–DNA binding events. The most widespread of these techniques is the genome-wide location analysis (GWLA), also known as chromatin immunoprecipitation coupled with microarray chip (ChIP–chip) or chromatin immunoprecipitation coupled with paired-end ditag sequencing (ChIP–PET) [73–76]. In GWLA, TR–DNA binding events are captured and frozen in a specific cellular state by in vivo crosslinking. Then the genomic DNA is fragmented and the TR of interest is isolated with a specific antibody, along with those genomic fragments bounded by the TR. After crosslinking reversal and protein digestion, the pulled-down DNA is labeled in a manner analogous to a cDNA microarray experiment, but hybridized to an oligo microarray chip whose content is directed toward regulatory regions rather than exons. These chips are composed of 20- to 70-mer probes tiling the entire genome or more likely, in the case of mammalian genomes, tiling promoters and intergenic regions. In the case of two-colored array platforms, the immunoprecipitated DNA can be compared to the total DNA input (the same genomic DNA but divided before the immunoprecipitation step) and this facilitates the normalization of the fluorescent signals [75].

In the case of ChIP–PET, the immunoprecipitated DNA is cloned into a DNA library and then converted into paired-end ditags (PETs). The PETs are concatenated and cloned into a ChIP–PET library for sequencing, and the number of ditags should be proportional to the original amount of immunoprecipitated DNA fragment and therefore enriched for those fragment bounded by the TR [76].

GWLAs are powerful techniques since they capture in vivo binding events in a high-throughput fashion and thus TF binding events can be compared across several cellular conditions, drug stimulation, developmental stage, etc. GWLAs also facilitate computational prediction of TFBSs since the experimentally identified TR binding regions drastically reduce the search space.

The general limitation of GWLAs is the high noise level (and so high false discovery rate) in the data due to several technical challenges in the experimental protocol. The amplification step required for the DNA labeling tends to reduce the enrichment of the immunoprecipitated DNA, and, most importantly, the quality and specificity of the antibody can change the efficiency of the immunoprecipitation. The antibody also introduces a limitation in the horizontal throughput of these techniques, meaning it is difficult, particularly in mammalian systems, to map the binding events of several TRs in parallel so the selection of the TRs is strongly biased toward the availability of a high-quality antibody. This problem is less accentuated in model organisms such as yeast in which all the TRs have been epitope-tagged using the tandem affinity purification (TAP) tag [77] and therefore the same anti-TAP antibody can be use to immunoprecipitate virtually all the TRs in an array of different conditions [78,79] (Table 1).

Computational modeling and inference of regulatory networks

Before the advent of high-throughput assays such as GWLA and CAGE, the most common methods to infer TR–DNA binding events were computational ones. In a network view we can draw inferred edges from a TR to a gene that bears a putative binding site(s) for the respective TR in its regulatory region. All these approaches have several intrinsic limitations, especially when applied to complex organisms such as humans, as the genome size correlates directly with the amount of noise in binding-site predictions. In addition, mammalian promoters are not well or easily defined because a promoter regulating a gene can reside a long distance away from the gene start and multiple promoter regions can contribute to the regulation of the same gene [80–83]. Taken together these factors make TFBS prediction a very difficult task and it may therefore suffer from problems of false positives and false negatives; yet there are several tricks (discussed below) that have been used successfully to reduce the noise in TR binding-site predictions. The advantage of TFBS prediction in the construction of TR–DNA edge libraries is that it can be done for virtually all the TRs for which a DNA binding motif has been defined.

Identification of regulatory DNA elements

Cross-species sequence comparisons, which rely upon the slow substitution rate of many categories of functional DNA relative to neutral sequence, have emerged as the preeminent means of identifying candidate *cis*-regulatory elements in mammalian genomes [84–89]. These studies involve sequence comparisons of human (or other mammal) genomic intervals to orthologous regions from organisms separated by varying evolutionary distances, ranging from primates to fish. The most important issue of comparative analysis is the choice of species, which depends on the goal being pursued. Previous theoretical studies [90,91] have shown that higher-resolution functional prediction at the level of a transcription-factor binding site (6–12 bp) is likely to require sequence from more than 10 mammals spread across the clade. In practice, Xie et al. performed a comparative analysis of the human, mouse, rat, and dog genomes to create a systematic catalog of common regulatory motifs in promoters and 3′ untranslated regions (3′ UTRs). The promoter analysis yielded 174 candidate motifs, including most previously known transcription factor binding sites and 105 new motifs. The 3′-UTR analysis yielded 106 motifs likely to be involved in posttranscriptional regulation. On the other hand, Pennachio made use of extreme evolutionary sequence conservation as a filter to identify putative enhancer activity of a large group of noncoding elements in the human genome that are conserved in human–pufferfish (*Takifugu* (Fugu) *rubripes*) or ultraconserved in human–mouse–rat. They tested 167 of these extremely conserved sequences in a transgenic mouse enhancer

assay and found that 45% of these sequences functioned reproducibly as tissue-specific enhancers of gene expression. By extrapolation, the authors estimated that there were at least an additional 5500 human–fish conserved noncoding sequences in the human genome with similar levels of constraint that are strong candidates for acting as gene enhancers. Sequencing of additional mammalian genomes [92] will incrementally facilitate the identification of large regulatory modules in the human genome. The advantage of the comparative genomic approach over current high-throughput technologies is that it can identify conserved regulatory modules.

With the advent of high-throughput technologies such as GWLA and CAGE, many types of functional DNA elements (TFBSs, basal promoters, enhancers, insulators, etc.) are now being experimentally mapped on the genome scale with a typical resolution of a few hundred nucleotides. Carninci, Nilsson, and Tegner, for example, have recently used CAGE-based TSS mapping to restrict the *cis*-regulatory elements search space and also to focus the analysis on only active (expressed) REs, thereby reducing the number of false positives.

Raw data from these high-throughput assays are noisy and typically do not have resolution at the single binding site level. Thus several new computational methods for DNA motif finding have been developed to take advantage of genome-wide location data. ChIP–chip binding *p* values were first used to rank and select high-confidence promoter sequences as inputs to existing motif discovery algorithms [93,94]. Recently, two groups introduced boosting strategies that take into account both binding and nonbinding sequences from ChIP–chip data during motif search [95,96].

In addition to narrowing the search space, GWLA data also provide quantitative information about the binding free energy of TR–DNA interaction. Two groups [97,98] have fitted a statistical–mechanical model of TR–DNA interactions to binding ratios of oligonucleotide probes on the microarray used in ChIP–chip, to infer DNA motif models for the TRs.

GWLA technologies have also been used to map chromatin modification states, such as histone modifications. In a recent study, Heintzman et al. [74] demonstrated that these epigenetic signatures could be used to identify promoters and enhancers in the human genome.

Identification of regulatory DNA modules

For genes in higher eukaryotes the binding of an individual TR is not sufficient to drive context-specific transcription. Rather, interaction and cooperation of several TRs are needed to affect gene expression at specific times and locations [99]. The DNA regions targeted by a group of TRs are usually clustered together and form so-called *cis*-regulatory modules. Compared to TR binding-site predictions, module predictions are more reliable, and methods have been developed along several lines. In the first category the promoters of a set of coregulated genes obtained from prior experiments are analyzed to identify overrepresented motif combinations likely to be responsible for the genes' coregulation [100–105]. Other approaches assume that the user provides a small set of TF–DNA binding motifs that are expected to co-occur in modules and identify genomic regions densely populated in putative sites for these TRs [106–110]. A third type of approach is based on the detection of statistically significant clusters of phylogenetically conserved TR binding sites [86,111,112].

Current databases for regulatory interactions

Large regulatory interaction datasets are now available for a variety of metazoan species (Table 2), including *Drosophila melanogaster, Caenorhabditis elegans,* and *Homo sapiens.* In light of these vast scientific resources made available through experimental and computational analyses, several databases storing interaction data are now in wide usage (Table 2). Most of these databases contain interaction data derived from both high-throughput analyses and small-scale experiments. In addition to being data warehouses, some of these databases have developed new methods for data exchange and visualization to facilitate the study of molecular interaction networks.

Integrating gene expression profiles with molecular interaction data to construct regulatory networks

Gene expression profiles describe the steady-state mRNA levels in the cell—the outcome of the regulatory network. Although network structure could in principle be inferred from only expression data, it would be a very challenging task because of the small sample sizes (number of genes greatly exceeds number of measurements per gene) and large amount of noise in expression profiles. However, the integration of large-scale physical interaction datasets with expression data provides a more direct route for reconstructing gene regulatory networks.

Different data sources have their own limitations. Currently, both gene expression and GWLA data are noisy. Differentially expressed genes from replicate microarray experiments typically overlap by 70–75% [113], whereas the overlap between replicate ChIP–chip experiments is even lower, usually less than 50%. Because gene expression and TR location data

Table 2
Databases of transcription regulatory and biomolecular physical interactions in metazoan

Type of interaction	Web address
Transcription regulatory	
TRANSFAC	www.gene-regulation.com/pub/databases.html#transfac
TRED	http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=home
VISTA Enhancer	http://genome.lbl.gov/vista/index.shtml
PReMod	http://genomequebec.mcgill.ca/PReMod
Biomolecular physical	
DIP	http://dip.doe-mbi.ucla.edu
BIND	www.binddb.org
HPRD	www.hprd.org
REACTOME	www.reactome.org
MINT	http://mint.bio.uniroma2.it/mint/Welcome.do
PDZBase	http://icb.med.cornell.edu/services/pdz/start
AfCS	www.signaling-gateway.org
IntAct	www.ebi.ac.uk/intact/site/index.jsf

provide complementary information, integration of the two data sources can emphasize the functional part of the network and thus make the inferred network more biologically relevant. This observation provides the rationale for several recently developed integrative approaches. Bar-Joseph and colleagues developed an iterative method in which coherent gene expression profile is used to include low-confidence but true ChIP–chip targets in regulatory modules [114]. Gao et al. [115] used multiple linear regressions to model gene expression ratios and TF binding to gene promoters, as measured with ChIP–chip. Similarly, Liao et al. [116] used two sets of linear equations to model simultaneously the regulatory strength of TFs on their DNA targets and the activity levels of the TF themselves. This is an improvement over the linear regression approach of Gao and colleagues, which could model only TF activity levels.

Most cellular functions are carried out by protein complexes, such as ribosome, spliceosome, and proteasome. So far little is known about how protein complexes are regulated. Protein interaction data could also be integrated with gene expression profiles to discover regulated protein complexes [117,118] and to study the regulatory dynamics of protein complexes [119].

Finally, gene expression data could also be integrated simultaneously with both protein–DNA (GWLA) and protein–protein interaction data (Fig. 3). Examples in this category include the jActiveModule [120] and the physical network [121] approach. jActiveModule, as the name suggests, identifies a set of genes that are differentially expressed under certain conditions and whose protein products physically interact. The physical network approach by Yeang et al. identifies causal physical pathways leading to differentially expressed genes.

Using integrated networks

A regulatory network, defined from several different data sources as outlined above, is useful in different ways. Clearly the network structure suggests novel mechanistic hypotheses

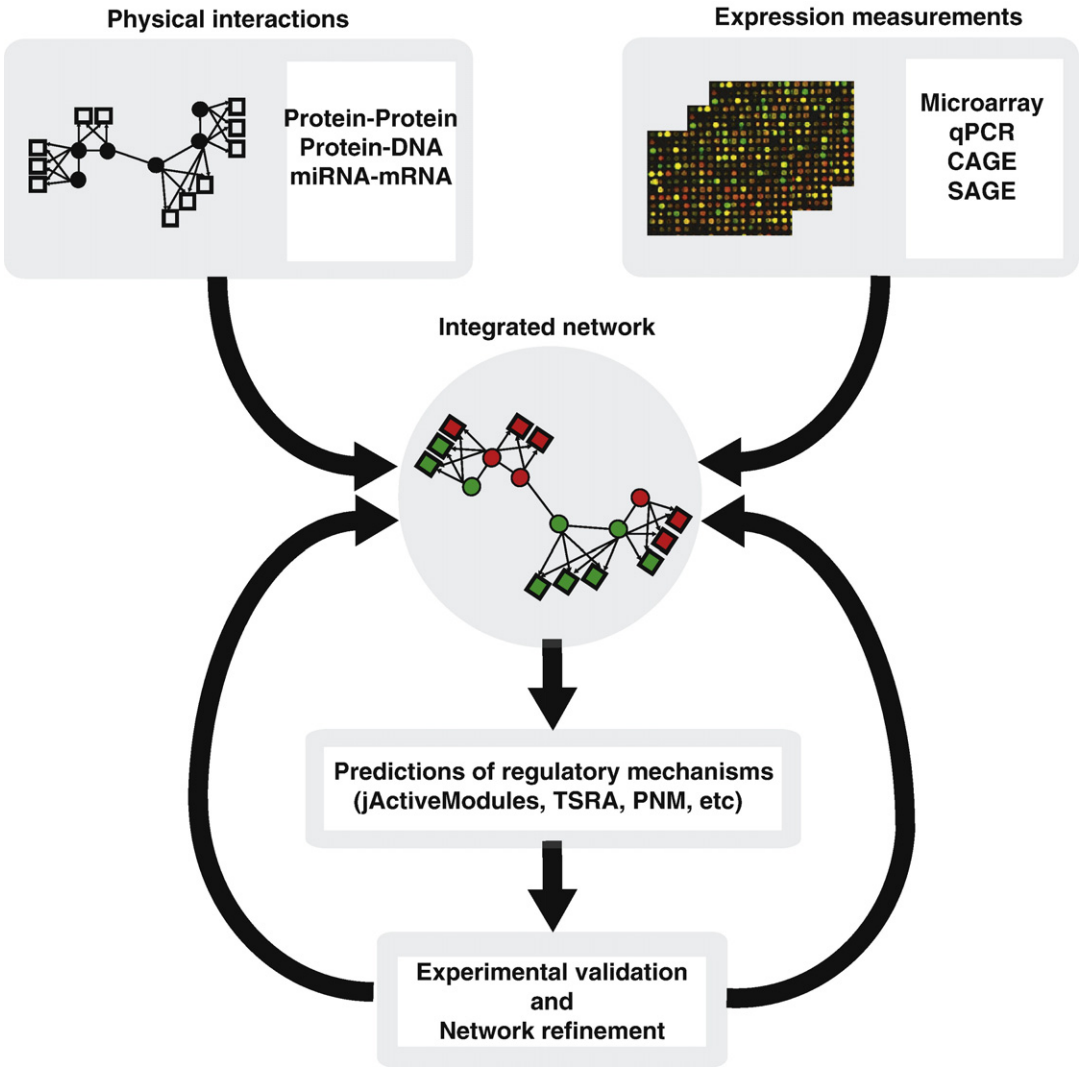

Fig. 3. An integrated approach to uncovering transcription regulatory networks. Representation of an integrated approach similar to the one used by Nilsson et al. [39], in which physical interactions (protein–protein, protein–DNA, etc.) and expression measurements (microarray, qPCR, CAGE, etc.) are combined to generate an integrated network model. Predictions of regulatory mechanisms, obtained from the network interrogation using bioinformatics tools such as jActiveModules, PNM, and TSRA [39,120,121], are first experimentally validated and then used to refine the original network model.

that must be experimentally tested as a final validation step. However, before this step is taken it is mandatory to consider that networks are condition-and state-dependent, that is, different parts of the network will be active under different conditions. For example, a cell that is exposed to a particular compound or a physiological condition such as stress will produce two different activity patterns. Therefore a static network defined from different data sources has to be evaluated and projected onto the specific condition of interest. Such a network projection can be performed in space (over different organs/tissues) and/or in time (in response to a stimulation for example).

Integrated approaches to studying tissue-and disease-specific networks

The tissue-specific mRNA expression patterns of a gene can offer important clues to its physiological function. Previously, Su et al. have generated a large compendium of gene expression profiles of 79 human and 61 mouse tissues [122]. Apart from normal tissues, a large number of disease-specific expression profiles (cancer, immune disorders, neurological diseases, etc.) have also been generated over the past decades. These expression datasets, which in essence constitute an activity-defined "fingerprint" of a disease state, provide unprecedented opportunities to study transcriptional regulatory networks in mammals. The key to success is to adopt a system to integrate these expression profiles with the static (interactome) network and phenotypic data. Recently, as a proof-of-concept, several groups have adopted this strategy to study regulatory networks in human diseases, including cancer [123,124], innate immunity [39,125], and inflammation [126]. For example Segal and coworkers used a large compendium of gene expression profiles to defined shared and cancer-subtype-specific modules [124]. Similarly, Paulsson and coworkers have defined metabolic interaction networks and used these static networks together with expression data sampled from different conditions to define disease states of relevance for diabetes and cardiovascular diseases [127].

Network dynamics

In addition to spatial, tissue, and disease-dependent activity, it is equally important to investigate system dynamics over a period

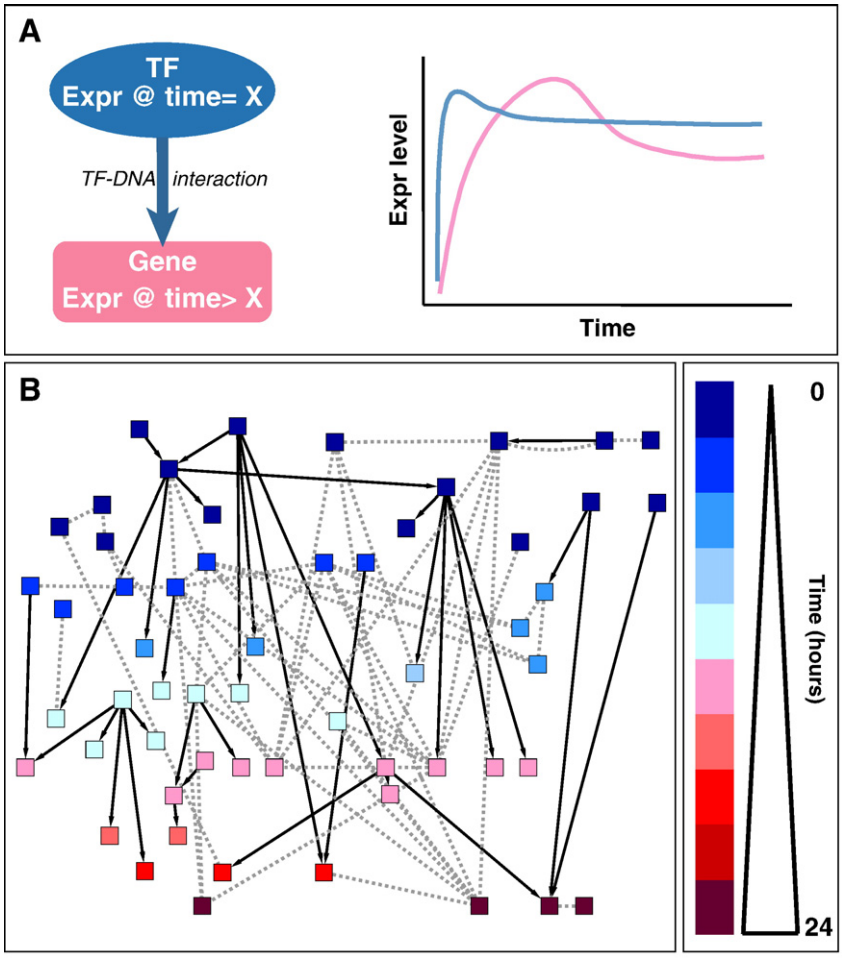

Fig. 4. Transcription regulatory network dynamics in macrophage activation by lipopolysaccharide (LPS), a bacterial endotoxin, obtained by applying the integrated approach developed by Nilsson and colleagues [39]. (A) The integrated network is parsed for time-shifted regulatory events involving TFs and their regulated genes. (B) Network model inferred using the TSR algorithm. Nodes represent TFs and are colored based on the time of maximum expression during the LPS time course (0 to 24 h poststimulation). Arrowed edges indicate TF–DNA binding events inferred as described by Nilsson and colleagues. Dotted edges are protein–protein interactions.

of time. This approach has a long tradition based on computational modeling of cellular processes [128]. Those purely computational attempts are severely limited by the current lack of information about kinetic parameters describing the network dynamics. In contrast, using a topological structure defined by integration of different data types offers a much simpler temporal analysis without the need for detailed information and the precise kinetics. Here the relevant parts of a static network can instead be extracted when the system (cell/populations/tissues/organs) produces a time-dependent response. For example, a human cell or a cell line can be stimulated to mimic a physiological response of interest. Calvano and coworkers [126] used manual curation to assemble different sources of data that defined a static inflammatory network in human. Next, they studied the blood leukocyte gene expression patterns at different time points (2, 4, 6, 9, 24 h) in human subjects receiving an inflammatory stimulus (bacterial endotoxin). The infusion of endotoxin essentially activated an innate immune response of brief duration. By extracting different parts of their inflammatory network at each time point the authors revealed an initial proinflammatory network and a subsequent counterregulatory network. A similar analysis of allergic disease has been presented by Benson and colleagues [129]. Nilsson and colleagues [39] used a promoter-based network as a backbone to study the LPS-induced response in bone-marrow-derived macrophages (Fig. 4). Again, by extracting the active part of the network at different time points, using an algorithm that searches the network for time-shifted regulatory events involving TFs and their regulated genes, the authors uncovered central pathways involved in the inflammatory response and new critical roles for TRs during the inflammatory response (Fig. 4). These particular examples demonstrate how integrative networks can be used to uncover time-dependent and disease-relevant responses, which provides a wealth of mechanistic hypotheses that can be subjected to subsequent experimental analysis.

Future challenges

A major emerging challenge of network biology is to compare and contrast biological networks systematically over different species, conditions, cell types, disease states, or points in time. For this purpose, methods are being developed to compare and contrast protein interaction networks to predict regulatory interactions [118,130] and to identify conserved interaction complexes and pathways [118]. Although most of the previous research has focused on protein interaction networks, many methods can be extended to compare transcriptional regulatory networks. An intriguing future application of these integrated networks may be to probe the functional role of single nucleotide polymorphisms, which are now rapidly being assembled for a range of diseases [131,132].

Finally, it is clear that data-integration methods need to be formulated in a proper statistical framework. For example, for each data type it is essential to develop statistical measures that indicate the reliability of the data, thus setting the stage for a weighted integrative method in which each data source contributes in proportion to its internal quality and relevance to the question at hand.

Acknowledgments

K.T. was funded by a postdoctoral fellowship from IBM Corp., through its Institutes of Innovation Program. J.T. was supported by Clinical Gene Networks AB and Linköping University. T.R. was supported by National Institute of Mental Health Grant 2P30MH062261-07.

References

[1] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.

[2] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.

[3] E.H. Davidson, D.H. Erwin, Gene regulatory networks and the evolution of animal body plans, Science 311 (2006) 796–800.

[4] E.H. Davidson, D.R. McClay, L. Hood, Regulatory gene networks and the properties of the developmental process, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 1475–1480.

[5] M. Levine, E.H. Davidson, Gene regulatory networks for development, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 4936–4942.

[6] S.C. Materna, E.H. Davidson, Logic of gene regulatory networks, Curr. Opin. Biotechnol. 18 (2007) 351–354.

[7] L. De Lellis, M.C. Curia, G.M. Aceto, S. Toracchio, G. Colucci, A. Russo, R. Mariani-Costantini, A. Cama, Analysis of extended genomic rearrangements in oncological research, Ann. Oncol. 18 (Suppl 6) (2007) vi173–vi178.

[8] S.J. Tebbutt, A. James, P.D. Pare, Single-nucleotide polymorphisms and lung disease: clinical implications, Chest 131 (2007) 1216–1223.

[9] V.E. Carlton, J.S. Ireland, F. Useche, M. Faham, Functional single nucleotide polymorphism-based association studies, Hum. Genomics 2 (2006) 391–402.

[10] T. Misteli, Beyond the sequence: cellular organization of genome function, Cell 128 (2007) 787–800.

[11] T. Cremer, M. Cremer, S. Dietzel, S. Muller, I. Solovei, S. Fakan, Chromosome territories—a functional nuclear landscape, Curr. Opin. Cell Biol. 18 (2006) 307–316.
[12] N. Gilbert, S. Gilchrist, W.A. Bickmore, Chromatin organization in the mammalian nucleus, Int. Rev. Cytol. 242 (2005) 283–336.
[13] T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells, Nat., Rev. Genet. 2 (2001) 292–301.
[14] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V.B. Bajic, S.E. Brenner, S. Batalov, A.R. Forrest, M. Zavolan, M.J. Davis, L.G. Wilming, V. Aidinis, J.E. Allen, A. Ambesi-Impiombato, R. Apweiler, R.N. Aturaliya, T.L. Bailey, M. Bansal, L. Baxter, K.W. Beisel, T. Bersano, H. Bono, A.M. Chalk, K.P. Chiu, V. Choudhary, A. Christoffels, D.R. Clutterbuck, M.L. Crowe, E. Dalla, B.P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C.F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T.R. Gingeras, T. Gojobori, R.E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T.K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S.P. Krishnan, A. Kruger, S.K. Kummerfeld, I.V. Kurochkin, L.F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, et al., The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563.
[15] S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C.C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K.C. Pang, J. Hallinan, J. Mattick, D.A. Hume, L. Lipovich, S. Batalov, P.G. Engstrom, Y. Mizuno, M.A. Faghihi, A. Sandelin, A.M. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, C. Wahlestedt, Antisense transcription in the mammalian transcriptome, Science 309 (2005) 1564–1566.
[16] T. Ravasi, H. Suzuki, K.C. Pang, S. Katayama, M. Furuno, R. Okunishi, S. Fukuda, K. Ru, M.C. Frith, M.M. Gongora, S.M. Grimmond, D.A. Hume, Y. Hayashizaki, J.S. Mattick, Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome, Genome Res. 16 (2006) 11–19.
[17] J. Tegner, J. Bjorkegren, Perturbations to uncover gene networks, Trends Genet. 23 (2007) 34–41.
[18] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, D. di Bernardo, How to infer gene networks from expression profiles, Mol. Syst. Biol. 3 (2007) 122.
[19] J. Tegner, M.K. Yeung, J. Hasty, J.J. Collins, Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5944–5949.
[20] T.S. Gardner, D. di Bernardo, D. Lorenz, J.J. Collins, Inferring genetic networks and identifying compound mode of action via expression profiling, Science 301 (2003) 102–105.
[21] C. Christensen, J. Thakar, R. Albert, Systems-level insights into cellular regulation: inferring, analysing, and modelling intracellular networks, IET Syst. Biol. 1 (2007) 61–77.
[22] W. Gesler, The uses of spatial analysis in medical geography: a review, Soc. Sci. Med. 23 (1986) 963–973.
[23] P. Grindrod, M. Kibble, Review of uses of network and graph theory concepts within proteomics, Expert. Rev. Proteomics 1 (2004) 229–238.
[24] M.J. Keeling, K.T. Eames, Networks and epidemic models, J. R. Soc. Interface 2 (2005) 295–307.
[25] R.U. Muller, M. Stead, J. Pach, The hippocampus as a cognitive graph, J. Gen. Physiol. 107 (1996) 663–694.
[26] M.E. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev., E Stat. Nonlin. Soft Matter Phys. 64 (2001).
[27] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.
[28] J. DeRisi, L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, J.M. Trent, Use of a cDNA microarray to analyse gene expression patterns in human cancer, Nat. Genet. 14 (1996) 457–460.
[29] R.D. Canales, Y. Luo, J.C. Willey, B. Austermiller, C.C. Barbacioru, C. Boysen, K. Hunkapiller, R.V. Jensen, C.R. Knight, K.Y. Lee, Y. Ma, B. Maqsodi, A. Papallo, E.H. Peters, K. Poulter, P.L. Ruppel, R.R. Samaha, L. Shi, W. Yang, L. Zhang, F.M. Goodsaid, Evaluation of DNA microarray results with quantitative gene expression platforms, Nat. Biotechnol. 24 (2006) 1115–1122.
[30] J. Quackenbush, Microarray analysis and tumor classification, N. Engl. J. Med. 354 (2006) 2463–2472.
[31] D.W. Lin, P.S. Nelson, Microarray analysis and tumor classification, N. Engl. J. Med. 355 (2006) 960 (author reply 960).
[32] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C. Holstege, I.F. Kim, V. Markowitz, J.C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, M. Vingron, Minimum information about a microarray experiment (MIAME)—toward standards for microarray data, Nat. Genet. 29 (2001) 365–371.
[33] V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, Serial analysis of gene expression, Science 270 (1995) 484–487.
[34] S. Saha, A.B. Sparks, C. Rago, V. Akmaev, C.J. Wang, B. Vogelstein, K.W. Kinzler, V.E. Velculescu, Using the transcriptome to annotate the genome, Nat. Biotechnol. 20 (2002) 508–512.
[35] P. Carninci, Tagging mammalian transcription complexity, Trends Genet. 22 (2006) 501–510.
[36] P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C.A. Semple, M.S. Taylor, P.G. Engstrom, M.C. Frith, A.R. Forrest, W.B. Alkema, S.L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S.M. Grimmond, C.A. Wells, V. Orlando, C. Wahlestedt, E.T. Liu, M. Harbers, J. Kawai, V.B. Bajic, D.A. Hume, Y. Hayashizaki, Genome-wide analysis of mammalian promoter architecture and evolution, Nat. Genet. 38 (2006) 626–635.
[37] R. Kodzius, M. Kojima, H. Nishiyori, M. Nakamura, S. Fukuda, M. Tagami, D. Sasaki, K. Imamura, C. Kai, M. Harbers, Y. Hayashizaki, P. Carninci, CAGE: cap analysis of gene expression, Nat. Methods 3 (2006) 211–222.
[38] J.B. Kim, G.J. Porreca, L. Song, S.C. Greenway, J.M. Gorham, G.M. Church, C.E. Seidman, J.G. Seidman, Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy, Science 316 (2007) 1481–1484.
[39] R. Nilsson, V.B. Bajic, H. Suzuki, D. di Bernardo, J. Bjorkegren, S. Katayama, J.F. Reid, M.J. Sweet, M. Gariboldi, P. Carninci, Y. Hayashizaki, D.A. Hume, J. Tegner, T. Ravasi, Transcriptional network dynamics in macrophage activation, Genomics 88 (2006) 133–142.
[40] J. Tegner, R. Nilsson, V.B. Bajic, J. Bjorkegren, T. Ravasi, Systems biology of innate immunity, Cell. Immunol. 244 (2006) 105–109.
[41] M.J. Holland, Transcript abundance in yeast varies over six orders of magnitude, J. Biol. Chem. 277 (2002) 14363–14366.
[42] J.V. Falvo, A.M. Uglialoro, B.M. Brinkman, M. Merika, B.S. Parekh, E.Y. Tsai, H.C. King, A.D. Morielli, E.G. Peralta, T. Maniatis, D. Thanos, A.E. Goldfeld, Stimulus-specific assembly of enhancer complexes on the tumor necrosis factor alpha gene promoter, Mol. Cell. Biol. 20 (2000) 2239–2247.
[43] T.K. Kim, T. Maniatis, The mechanism of transcriptional synergy of an in vitro assembled interferon-beta enhanceosome, Mol. Cell 1 (1997) 119–129.
[44] M. Bellorini, D.K. Lee, J.C. Dantonel, K. Zemzoumi, R.G. Roeder, L. Tora, R. Mantovani, CCAAT binding NF-Y–TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues, Nucleic Acids Res. 25 (1997) 2174–2181.
[45] A. Hoffmann, T. Oelgeschlager, R.G. Roeder, Considerations of transcriptional control mechanisms: do TFIID–core promoter complexes recapitulate nucleosome-like functions? Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 8928–8935.

[46] A.B. Lassar, P.L. Martin, R.G. Roeder, Transcription of class III genes: formation of preinitiation complexes, Science 222 (1983) 740–748.
[47] T. Oelgeschlager, C.M. Chiang, R.G. Roeder, Topology and reorganization of a human TFIID–promoter complex, Nature 382 (1996) 735–738.
[48] M.W. Van Dyke, M. Sawadogo, R.G. Roeder, Stability of transcription complexes on class II genes, Mol. Cell. Biol. 9 (1989) 342–344.
[49] S.C. Tucker, R. Wisdom, Site-specific heterodimerization by paired class homeodomain proteins mediates selective transcriptional responses, J. Biol. Chem. 274 (1999) 32325–32332.
[50] J.A. Armstrong, B.M. Emerson, Transcription of chromatin: these are complex times, Curr. Opin. Genet. Dev. 8 (1998) 165–172.
[51] M. Gaestel, Molecular chaperones in signal transduction, Handbk. Exp. Pharmacol. (2006) 93–109.
[52] R.X. Luo, D.C. Dean, Chromatin remodeling and transcriptional regulation, J. Natl. Cancer Inst. 91 (1999) 1288–1294.
[53] T. Tsukiyama, C. Wu, Chromatin remodeling and transcription, Curr. Opin. Genet. Dev. 7 (1997) 182–191.
[54] A. Dvir, J.W. Conaway, R.C. Conaway, Mechanism of transcription initiation and promoter escape by RNA polymerase II, Curr. Opin. Genet. Dev. 11 (2001) 209–214.
[55] M. Kozak, Initiation of translation in prokaryotes and eukaryotes, Gene 234 (1999) 187–208.
[56] M. Kunzler, C. Springer, G.H. Braus, The transcriptional apparatus required for mRNA encoding genes in the yeast Saccharomyces cerevisiae emerges from a jigsaw puzzle of transcription factors, FEMS Microbiol. Rev. 19 (1996) 117–136.
[57] C.A. Spencer, M. Groudine, Transcription elongation and eukaryotic gene regulation, Oncogene 5 (1990) 777–785.
[58] T. Tamura, Y. Makino, T. Kishimoto, Regulation of gene expression and recent advance on transcription studies, Nippon Rinsho 53 (1995) 1033–1047.
[59] J.K. Joung, E.I. Ramm, C.O. Pabo, A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 7382–7387.
[60] S.A. Johnston, J.E. Hopper, Isolation of the yeast regulatory gene GAL4 and analysis of its dosage effects on the galactose/melibiose regulon, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 6971–6975.
[61] H. Suzuki, Y. Fukunishi, I. Kagawa, R. Saito, H. Oda, T. Endo, S. Kondo, H. Bono, Y. Okazaki, Y. Hayashizaki, Protein–protein interaction panel using mouse full-length cDNAs, Genome Res. 11 (2001) 1758–1765.
[62] H. Suzuki, R. Saito, M. Kanamori, C. Kai, C. Schonbach, T. Nagashima, J. Hosaka, Y. Hayashizaki, The mammalian protein–protein interaction database and its viewing system that is linked to the main FANTOM2 viewer, Genome Res. 13 (2003) 1534–1541.
[63] R.M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M.D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y.V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J.P. Lambert, H.S. Duewel, I.I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.L. Adams, M.F. Moran, G.B. Morin, T. Topaloglou, D. Figeys, Large-scale mapping of human protein–protein interactions by mass spectrometry, Mol. Syst. Biol. 3 (2007) 89.
[64] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (2003) 198–207.
[65] A.C. Gingras, R. Aebersold, B. Raught, Advances in protein complex analysis using mass spectrometry, J. Physiol. 563 (2005) 11–21.
[66] B. Blagoev, I. Kratchmarova, S.E. Ong, M. Nielsen, L.J. Foster, M. Mann, A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling, Nat. Biotechnol. 21 (2003) 315–318.
[67] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, M. Tyers, Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry, Nature 415 (2002) 180–183.
[68] M. Mann, R.C. Hendrickson, A. Pandey, Analysis of proteins and proteomes by mass spectrometry, Ann. Rev. Biochem. 70 (2001) 437–473.
[69] D. Lin, D.L. Tabb, J.R. Yates III, Large-scale protein identification using mass spectrometry, Biochim. Biophys. Acta 1646 (2003) 1–10.
[70] S. Mathivanan, B. Periaswamy, T.K. Gandhi, K. Kandasamy, S. Suresh, R. Mohmood, Y.L. Ramachandra, A. Pandey, An evaluation of human protein–protein interaction data in the public domain, BMC Bioinform. 7 (Suppl 5) (2006) S19.
[71] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D.S. Goldberg, L.V. Zhang, S.L. Wong, G. Franklin, S. Li, J.S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R.S. Sikorski, J. Vandenhaute, H.Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M.E. Cusick, D.E. Hill, F.P. Roth, M. Vidal, Towards a proteome-scale map of the human protein–protein interaction network, Nature 437 (2005) 1173–1178.
[72] G.T. Hart, A.K. Ramani, E.M. Marcotte, How complete are current yeast and human protein-interaction networks? Genome Biol. 7 (2006) 120.
[73] G.M. Euskirchen, J.S. Rozowsky, C.L. Wei, W.H. Lee, Z.D. Zhang, S. Hartman, O. Emanuelsson, V. Stolc, S. Weissman, M.B. Gerstein, Y. Ruan, M. Snyder, Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies, Genome Res. 17 (2007) 898–909.
[74] N.D. Heintzman, R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, B. Ren, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, Nat. Genet. 39 (2007) 311–318.
[75] B. Ren, F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T.L. Volkert, C.J. Wilson, S.P. Bell, R.A. Young, Genome-wide location and function of DNA binding proteins, Science 290 (2000) 2306–2309.
[76] C.L. Wei, Q. Wu, V.B. Vega, K.P. Chiu, P. Ng, T. Zhang, A. Shahab, H.C. Yong, Y. Fu, Z. Weng, J. Liu, X.D. Zhao, J.L. Chew, Y.L. Lee, V.A. Kuznetsov, W.K. Sung, L.D. Miller, B. Lim, E.T. Liu, Q. Yu, H.H. Ng, Y. Ruan, A global map of p53 transcription-factor binding sites in the human genome, Cell 124 (2006) 207–219.
[77] S. Ghaemmaghami, W.K. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O'Shea, J.S. Weissman, Global analysis of protein expression in yeast, Nature 425 (2003) 737–741.
[78] C.T. Harbison, D.B. Gordon, T.I. Lee, N.J. Rinaldi, K.D. Macisaac, T.W. Danford, N.M. Hannett, J.B. Tagne, D.B. Reynolds, J. Yoo, E.G. Jennings, J. Zeitlinger, D.K. Pokholok, M. Kellis, P.A. Rolfe, K.T. Takusagawa, E.S. Lander, D.K. Gifford, E. Fraenkel, R.A. Young, Transcriptional regulatory code of a eukaryotic genome, Nature 431 (2004) 99–104.
[79] T.I. Lee, N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, R.A. Young, Transcriptional regulatory networks in Saccharomyces cerevisiae, Science 298 (2002) 799–804.
[80] C.M. Gorman, Mammalian cell expression, Curr. Opin. Biotechnol. 1 (1990) 36–47.
[81] S. Majumder, M.L. DePamphilis, A unique role for enhancers is revealed during early mouse development, Bioessays 17 (1995) 879–889.
[82] F. Muller, P. Blader, U. Strahle, Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements, Bioessays 24 (2002) 564–572.
[83] L.C. Myers, R.D. Kornberg, Mediator of transcriptional regulation, Ann. Rev. Biochem. 69 (2000) 729–749.
[84] L.A. Pennacchio, E.M. Rubin, Genomic strategies to identify mammalian regulatory sequences, Nat., Rev. Genet. 2 (2001) 100–109.
[85] N. Ahituv, E.M. Rubin, M.A. Nobrega, Exploiting human–fish genome comparisons for deciphering gene regulation, Hum. Mol. Genet. 13 (Spec. No. 2) (2004) R261–R266.
[86] D.C. King, J. Taylor, L. Elnitski, F. Chiaromonte, W. Miller, R.C. Hardison, Evaluation of regulatory potential and conservation scores

for detecting cis-regulatory modules in aligned mammalian genome sequences, Genome Res. 15 (2005) 1051–1060.
[87] A. Woolfe, M. Goodson, D.K. Goode, P. Snell, G.K. McEwen, T. Vavouri, S.F. Smith, P. North, H. Callaway, K. Kelly, K. Walter, I. Abnizova, W. Gilks, Y.J. Edwards, J.E. Cooke, G. Elgar, Highly conserved non-coding sequences are associated with vertebrate development, PLoS Biol. 3 (2005) e7.
[88] X. Xie, J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, M. Kellis, Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals, Nature 434 (2005) 338–345.
[89] S. Prabhakar, F. Poulin, M. Shoukry, V. Afzal, E.M. Rubin, O. Couronne, L.A. Pennacchio, Close sequence comparisons are sufficient to identify human cis-regulatory elements, Genome Res. 16 (2006) 855–863.
[90] G.M. Cooper, E.A. Stone, G. Asimenos, E.D. Green, S. Batzoglou, A. Sidow, Distribution and intensity of constraint in mammalian genomic sequence, Genome Res. 15 (2005) 901–913.
[91] S.R. Eddy, A model of the statistical power of comparative genome sequence analysis, PLoS Biol. 3 (2005) e10.
[92] E.H. Margulies, J.P. Vinson, W. Miller, D.B. Jaffe, K. Lindblad-Toh, J.L. Chang, E.D. Green, E.S. Lander, J.C. Mullikin, M. Clamp, An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 4795–4800.
[93] X.S. Liu, D.L. Brutlag, J.S. Liu, An algorithm for finding protein–DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments, Nat. Biotechnol. 20 (2002) 835–839.
[94] E.M. Conlon, X.S. Liu, J.D. Lieb, J.S. Liu, Integrating regulatory motif discovery and genome-wide expression analysis, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 3339–3344.
[95] P. Hong, X.S. Liu, Q. Zhou, X. Lu, J.S. Liu, W.H. Wong, A boosting approach for motif modeling using ChIP–chip data, Bioinformatics 21 (2005) 2636–2643.
[96] L.Y. Wang, M. Snyder, M. Gerstein, BoCaTFBS: a boosted cascade learner to refine the binding sites suggested by ChIP–chip experiments, Genome Biol. 7 (2006) R102.
[97] B.C. Foat, A.V. Morozov, H.J. Bussemaker, Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE, Bioinformatics 22 (2006) e141–e149.
[98] A. Tanay, Extensive low-affinity transcriptional interactions in the yeast genome, Genome Res. 16 (2006) 962–972.
[99] N.M. Luscombe, M.M. Babu, H. Yu, M. Snyder, S.A. Teichmann, M. Gerstein, Genomic analysis of regulatory network dynamics reveals large topological changes, Nature 431 (2004) 308–312.
[100] Q. Zhou, W.H. Wong, CisModule: de novo discovery of cis-regulatory modules by hierarchical mixture modeling, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 12114–12119.
[101] W.W. Wasserman, J.W. Fickett, Identification of regulatory regions which confer muscle-specific gene expression, J. Mol. Biol. 278 (1998) 167–181.
[102] S. Aerts, P. Van Loo, Y. Moreau, B. De Moor, A genetic algorithm for the detection of new cis-regulatory modules in sets of coregulated genes, Bioinformatics 20 (2004) 1974–1976.
[103] E. Segal, R. Sharan, A discriminative model for identifying spatial cis-regulatory modules, J. Comput. Biol. 12 (2005) 822–834.
[104] W. Thompson, M.J. Palumbo, W.W. Wasserman, J.S. Liu, C.E. Lawrence, Decoding human regulatory circuits, Genome Res. 14 (2004) 1967–1974.
[105] M. Gupta, J.S. Liu, De novo cis-regulatory module elicitation for eukaryotic genomes, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 7079–7084.
[106] O. Johansson, W. Alkema, W.W. Wasserman, J. Lagergren, Identification of functional clusters of transcription factor binding motifs in genome sequences: the MSCAN algorithm, Bioinformatics 19 (Suppl. 1) (2003) i169–i176.
[107] T.L. Bailey, W.S. Noble, Searching for statistically significant regulatory modules, Bioinformatics 19 (Suppl. 2) (2003) II16–II25.
[108] M.C. Frith, M.C. Li, Z. Weng, Cluster-Buster: finding dense clusters of motifs in DNA sequences, Nucleic Acids Res. 31 (2003) 3666–3668.
[109] S. Sinha, E. van Nimwegen, E.D. Siggia, A probabilistic method to detect regulatory modules, Bioinformatics 19 Suppl. (1) (2003) i292–i301.
[110] W.B. Alkema, O. Johansson, J. Lagergren, W.W. Wasserman, MSCAN: identification of functional clusters of transcription factor binding sites, Nucleic Acids Res. 32 (2004) W195–W198.
[111] D. Kolbe, J. Taylor, L. Elnitski, P. Eswara, J. Li, W. Miller, R. Hardison, F. Chiaromonte, Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat, Genome Res. 14 (2004) 700–707.
[112] M. Blanchette, A.R. Bataille, X. Chen, C. Poitras, J. Laganiere, C. Lefebvre, G. Deblois, V. Giguere, V. Ferretti, D. Bergeron, B. Coulombe, F. Robert, Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression, Genome Res. 16 (2006) 656–668.
[113] L. Shi, L.H. Reid, W.D. Jones, R. Shippy, J.A. Warrington, S.C. Baker, P.J. Collins, F. de Longueville, E.S. Kawasaki, K.Y. Lee, Y. Luo, Y.A. Sun, J.C. Willey, R.A. Setterquist, G.M. Fischer, W. Tong, Y.P. Dragan, D.J. Dix, F.W. Frueh, F.M. Goodsaid, D. Herman, R.V. Jensen, C.D. Johnson, E.K. Lobenhofer, R.K. Puri, U. Schrf, J. Thierry-Mieg, C. Wang, M. Wilson, P.K. Wolber, L. Zhang, S. Amur, W. Bao, C.C. Barbacioru, A.B. Lucas, V. Bertholet, C. Boysen, B. Bromley, D. Brown, A. Brunner, R. Canales, X.M. Cao, T.A. Cebula, J.J. Chen, J. Cheng, T.M. Chu, E. Chudin, J. Corson, J.C. Corton, L.J. Croner, C. Davies, T.S. Davison, G. Delenstarr, X. Deng, D. Dorris, A.C. Eklund, X.H. Fan, H. Fang, S. Fulmer-Smentek, J.C. Fuscoe, K. Gallagher, W. Ge, L. Guo, X. Guo, J. Hager, P.K. Haje, J. Han, T. Han, H.C. Harbottle, S.C. Harris, E. Hatchwell, C.A. Hauser, S. Hester, H. Hong, P. Hurban, S.A. Jackson, H. Ji, C.R. Knight, W.P. Kuo, J.E. LeClerc, S. Levy, Q.Z. Li, C. Liu, Y. Liu, M.J. Lombardi, Y. Ma, S.R. Magnuson, B. Maqsodi, T. McDaniel, N. Mei, O. Myklebost, B. Ning, N. Novoradovskaya, M.S. Orr, T.W. Osborn, A. Papallo, T.A. Patterson, R.G. Perkins, E.H. Peters, R. Peterson, et al., The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements, Nat. Biotechnol. 24 (2006) 1151–1161.
[114] Z. Bar-Joseph, G.K. Gerber, T.I. Lee, N.J. Rinaldi, J.Y. Yoo, F. Robert, D.B. Gordon, E. Fraenkel, T.S. Jaakkola, R.A. Young, D.K. Gifford, Computational discovery of gene modules and regulatory networks, Nat. Biotechnol. 21 (2003) 1337–1342.
[115] F. Gao, B.C. Foat, H.J. Bussemaker, Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data, BMC Bioinformatics 5 (2004) 31.
[116] J.C. Liao, R. Boscolo, Y.L. Yang, L.M. Tran, C. Sabatti, V.P. Roychowdhury, Network component analysis: reconstruction of regulatory signals in biological systems, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 15522–15527.
[117] N. Simonis, J. van Helden, G.N. Cohen, S.J. Wodak, Transcriptional regulation of protein complexes in yeast, Genome Biol. 5 (2004) R33.
[118] K. Tan, T. Shlomi, H. Feizi, T. Ideker, R. Sharan, Transcriptional regulation of protein complexes within and across species, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1283–1288.
[119] U. de Lichtenberg, L.J. Jensen, S. Brunak, P. Bork, Dynamic complex formation during the yeast cell cycle, Science 307 (2005) 724–727.
[120] T. Ideker, O. Ozier, B. Schwikowski, A.F. Siegel, Discovering regulatory and signalling circuits in molecular interaction networks, Bioinformatics 18 (Suppl. 1) (2002) S233–S240.
[121] C.H. Yeang, T. Ideker, T. Jaakkola, Physical network models, J. Comput. Biol. 11 (2004) 243–262.
[122] A.I. Su, T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M.P. Cooke, J.R. Walker, J.B. Hogenesch, A gene atlas of the mouse and human protein-encoding transcriptomes, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 6062–6067.
[123] S.A. Tomlins, R. Mehra, D.R. Rhodes, X. Cao, L. Wang, S.M. Dhanasekaran, S. Kalyana-Sundaram, J.T. Wei, M.A. Rubin, K.J. Pienta, R.B. Shah, A.M. Chinnaiyan, Integrative molecular concept modeling of prostate cancer progression, Nat. Genet. 39 (2007) 41–51.
[124] E. Segal, N. Friedman, N. Kaminski, A. Regev, D. Koller, From signatures to models: understanding cancer using microarrays, Nat. Genet. 37 (2005) S38–S45 (Suppl).
[125] M. Gilchrist, V. Thorsson, B. Li, A.G. Rust, M. Korb, K. Kennedy, T. Hai, H. Bolouri, A. Aderem, Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4, Nature 441 (2006) 173–178.

[126] S.E. Calvano, W. Xiao, D.R. Richards, R.M. Felciano, H.V. Baker, R.J. Cho, R.O. Chen, B.H. Brownstein, J.P. Cobb, S.K. Tschoeke, C. Miller-Graziano, L.L. Moldawer, M.N. Mindrinos, R.W. Davis, R.G. Tompkins, S.F. Lowry, A network-based analysis of systemic inflammation in humans, Nature 437 (2005) 1032–1037.

[127] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, B.O. Palsson, Global reconstruction of the human metabolic network based on genomic and bibliomic data, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1777–1782.

[128] A. Ma'ayan, R.D. Blitzer, R. Iyengar, Toward predictive models of mammalian cells, Annu. Rev. Biophys. Biomol. Struct. 34 (2005) 319–349.

[129] M. Benson, L. Carlsson, G. Guillot, M. Jernas, M.A. Langston, M. Rudemo, B. Andersson, A network-based analysis of allergen-challenged $CD4^{+}$T cells from patients with allergic rhinitis, Genes Immun. 7 (2006) 514–521.

[130] H. Yu, N.M. Luscombe, H.X. Lu, X. Zhu, Y. Xia, J.D. Han, N. Bertin, S. Chung, M. Vidal, M. Gerstein, Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs, Genome Res. 14 (2004) 1107–1118.

[131] G.S. Sellick, L.R. Goldin, R.W. Wild, S.L. Slager, L. Ressenti, S.S. Strom, M.J. Dyer, F.R. Mauro, G.E. Marti, S. Fuller, M. Lyttelton, T.J. Kipps, M.J. Keating, T.G. Call, D. Catovsky, N. Caporaso, R.S. Houlston, A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia, Blood 110 (9) (Nov 1) (2007) 3326–3333.

[132] R. Saxena, B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H.N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E.K. Speliotes, M.R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S.B. Gabriel, G.W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke, S. Purcell, Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels, Science 316 (2007) 1331–1336.